



Physiographic and climatic controls on regional groundwater dynamics

Downloaded from: <https://research.chalmers.se>, 2023-05-05 12:28 UTC

Citation for the original published paper (version of record):

Haaf, E., Giese, M., Heudorfer, B. et al (2020). Physiographic and climatic controls on regional groundwater dynamics. *Water Resources Research*, 56(10).
<http://dx.doi.org/10.1029/2019WR026545>

N.B. When citing this work, cite the original published paper.

Water Resources Research



RESEARCH ARTICLE

10.1029/2019WR026545

Key Points:

- Important physiographic and climatic controls on groundwater dynamics in Southern Germany are quantified
- Empirical models are successfully constructed to explain groundwater dynamics based on climatic and physiographic system characteristics
- The constructed empirical models are generally consistent with physical process understanding

Supporting Information:

- Supporting Information S1

Correspondence to:

E. Haaf,
ezra.haaf@chalmers.se

Citation:

Haaf, E., Giese, M., Heudorfer, B., Stahl, K., & Barthel, R. (2020). Physiographic and climatic controls on regional groundwater dynamics. *Water Resources Research*, 56, e2019WR026545. <https://doi.org/10.1029/2019WR026545>

Received 16 OCT 2019

Accepted 27 SEP 2020

Accepted article online 2 OCT 2020

Author Contributions:

Conceptualization: Ezra Haaf




Formal analysis: Ezra Haaf

Methodology: Ezra Haaf, Markus Giese, Benedikt Heudorfer

Writing - original draft: Ezra Haaf

Writing - review & editing: Ezra Haaf, Markus Giese, Benedikt Heudorfer, Kerstin Stahl, Roland Barthel

Physiographic and Climatic Controls on Regional Groundwater Dynamics

Ezra Haaf^{1,2} , Markus Giese¹ , Benedikt Heudorfer³ , Kerstin Stahl³ , and Roland Barthel¹ 

¹Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden, ²Now at Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden, ³Faculty of Environment and Natural Resources, University of Freiburg, Breisgau, Germany

Abstract The main goal of this study is to explore whether the ideas established by surface water hydrologists in the context of predictions in ungauged basins can be useful in hydrogeology. The concrete question is whether it is possible to create predictive models for groundwater systems with no or few observations based on knowledge derived from similar groundwater systems which are well observed. To do so, this study analyzes the relationship between temporal dynamics of groundwater levels and climatic and physiographic characteristics. The analysis is based on data from 341 wells in Southern Germany with 10-year daily groundwater hydrographs. Observation wells are used in confined and unconfined sand and gravel aquifers from narrow mountainous valleys as well as more extensive lowland alluvial aquifers. Groundwater dynamics at each location are summarized with 46 indices describing features of groundwater hydrographs. Besides borehole log-derived geologic information, local and regional morphologic characteristics as well as topography-derived boundary and climatic descriptors were derived for each well. Regression relationships were established by mining the data for associations between dynamics and descriptors with forward stepwise regression at a confidence level >95%. The most important predictors are geology and boundary conditions and, secondarily, climate, as well as some topographic features, such as regional convergence. The multiple regression models are in general agreement with process understanding linked to groundwater dynamics in unconfined aquifers. This systematic investigation suggests that statistical regionalization of groundwater dynamics in ungauged aquifers based on map-derived physiographic and climatic controls may be feasible.

1. Introduction

Hydrogeological investigations and in particular groundwater resource assessment are strongly reliant on the dynamics of groundwater levels. However, records of groundwater levels are often scarce and unevenly distributed in space and time. This irregularity of measurements combined with a high variability of hydrogeological systems with heterogeneous properties and uncertain inputs and driving processes leads to the need for systematic methods for prediction of groundwater in poorly observed (ungauged) groundwater systems. The present study is part of a larger research effort where methods of comparative regional analysis are investigated to estimate groundwater level dynamics at ungauged sites based on similarity of groundwater system response and climatic and nonclimatic characteristics. Understanding which of these characteristics can be linked to groundwater dynamics is a first step to improve estimates of storage change and thus prediction of groundwater availability in ungauged aquifers. This paper investigates these links and their strength and discusses them in context of literature studying the relation between physiography and exhibited groundwater dynamics. Groundwater level changes (dynamics) are driven by recharge and discharge processes, when groundwater systems are unaffected by direct human impacts (such as pumping or infiltration). Recharge and discharge are further controlled by complex interactions between climatic and geologic, and topographic and hydrologic boundary conditions (Cuthbert, 2014; Winter, 2001). These factors lead to specific hydrogeological settings that dictate the groundwater dynamics (Knutsson & Fagerlind, 1977; Tóth, 1970; Zecharias & Brutsaert, 1988). Studies investigating these relationships so far have been primarily concerned with steady state regional flow characterization using numerical models with synthetic data (Freeze & Witherspoon, 1967; Gleeson & Manning, 2008; Tóth, 1963; Winter, 1978). Quantification of regional groundwater flow, however, do not directly resolve the question of controls of

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

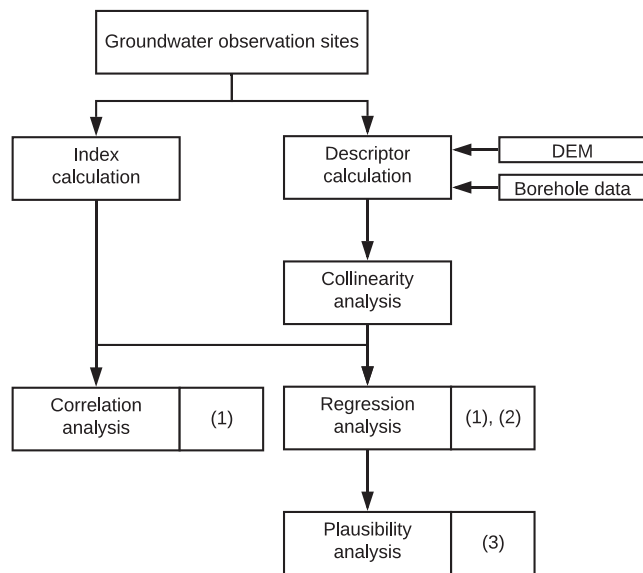


Figure 1. Workflow of analysis. Research questions in parentheses.

temporal dynamics of groundwater levels within an aquifer. When investigating these relations, researchers have often resorted to classification based on similarity of records of groundwater level time series (Haaf & Barthel, 2018; Lafare et al., 2015; Machiwal & Singh, 2015; Winter et al., 2000) and/or aquifer properties (Allen et al., 2010; Barthel, 2011). The former approaches have the disadvantage of being dependent on a classification method. The latter lead to limited transferability of the classification (Barthel, 2011; Haaf & Barthel, 2018; Heudorfer et al., 2019). Other researchers who have examined controls on groundwater dynamics have singled out certain aspects of the groundwater regime and investigated its controls. Cuthbert (2014), for example, described the relationship between groundwater recession and aquifer characteristics. Rinderer et al. (2016, 2019) searched for topographic controls on event-scale groundwater table responses. Boutt (2017) studied long-term observational records, linking summary statistics of groundwater hydrographs to drainage boundaries.

The present study takes advantage of the concept of hydrogeological similarity: Similarity in temporal groundwater dynamics is due to similar basin characteristics given similar forcing. This line of thinking has a long tradition within hydrology, as established by Falkenmark

et al. (1989) and expanded upon during the previous International Association of Hydrological Sciences decade (2003–2013) with the aim to facilitate regionalization and predictions in ungauged basins (PUB) (Blöschl et al., 2013). Unlike in surface water, similarity-based studies trying to link physical controls to groundwater dynamics are still limited in number (Boutt, 2017). Further, the studies generally do not cover many different types of generally available or map-derivable characteristics, such as topographic, hydrogeologic, and climatic controls. Additionally, none of these studies examine the entire range of dynamics aspects, i.e., event-scale, longer-structural, and distribution-based dynamics features, as identified in the classification of regime components by Heudorfer et al. (2019). Therefore, there is a need for a more exhaustive study with regard to the information retained in groundwater level time series to link dynamics to site characteristics. The purpose of this paper is therefore to systematically explore salient controls on groundwater dynamics in unconsolidated systems by using a comprehensive data set from Bavaria, Southern Germany, covering 50,000 km². This study is driven by three main research questions:

1. What are the most relevant descriptors to link climatic and physiographic properties to dynamics of groundwater time series?
2. To what extent can empirical models based on lumped site descriptors of physiography and climate explain groundwater dynamics?
3. Do these empirical relationships align with physical process understanding?

To answer the three research questions above, the following workflow was adapted (Figure 1). Groundwater observation wells were selected according to data availability and quality criteria (section 2.1). At each site, statistical aggregates (indices) of the groundwater level time series (section 2.1) and site descriptors (section 2.2) were calculated. Collinear site descriptors are eliminated before answering research question (1) with correlation analysis and multiple regression analysis to construct empirical models (research question (2)) and assess plausibility of constructed models (both section 2.3).

2. Data and Methods

2.1. Study Area and Groundwater Level Data

The study area comprises large parts of the upper Danube catchment in Bavaria, Southern Germany (Figure 2). The area of interest can be divided into three distinct areas: the Bavarian Alps to the South (ALP), the Alpine Foreland (AFO), and the Molasse basin (MOL). The climate is humid with an average rainfall of 600–1,800 mm/annum and mean annual temperature of −3.5°C to 10.2°C. The relatively large temperature and precipitation ranges are attributed to the topographic relief, from 2,600 m a.s.l. in the ALP

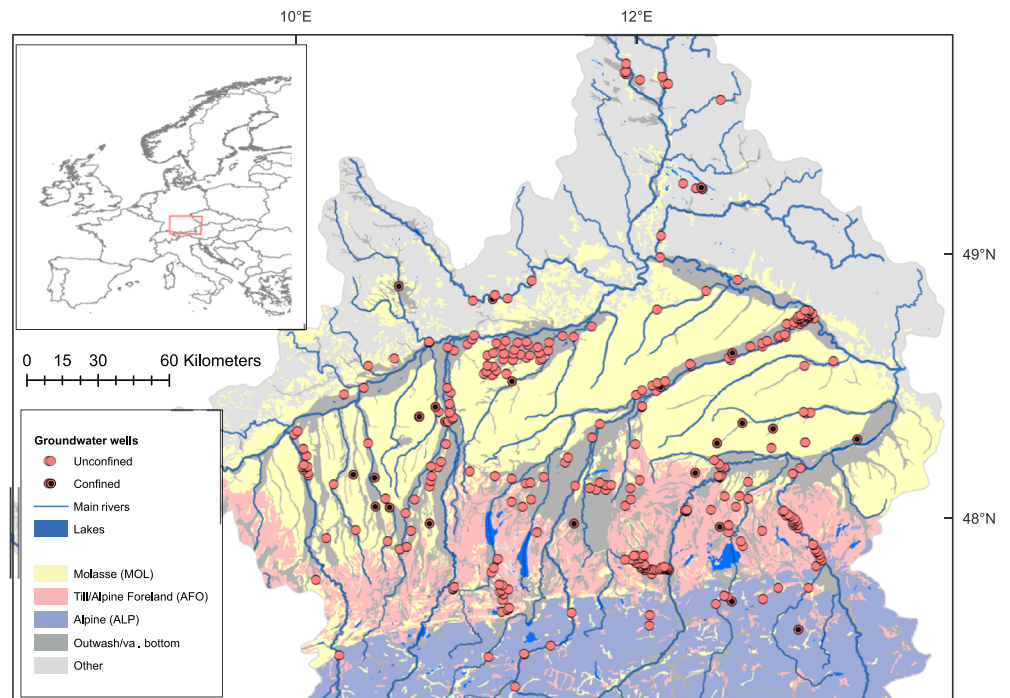


Figure 2. Spatial extent of the study, showing the observation well locations.

decreasing to around 800 m a.s.l. in the AFO, reaching as low as 300 m a.s.l. in the MOL toward the Danube, before rising to 1,000 m a.s.l. in the north. From a large data set within this area, a set of groundwater time series were selected based on availability of geological information and absence of patterns of direct anthropogenic impact (for a more detailed explanation, refer to Heudorfer et al., 2019). In this study, continuous time series were selected with daily measurements of more than 10 years and less than 1% missing data. The measurement periods are variable and do not necessarily cover the same period, but measurements cover the years between 1982 and 2017. A further selection criterion was availability of stratigraphic data at the well location from bore logs. The 341 observation wells used in this study are located mainly in Quaternary fluvial and glacio-fluvial sediments within the pre-alpine Molasse basin. The majority of the wells tap shallow Quaternary sediments in river valleys and fluvial sand and gravel deposits, while wells located in the deeper Tertiary sediments are fewer in number (Figure 2). For preprocessing, the daily groundwater hydrographs (D) were smoothed with a low-pass frequency filter to reduce noise associated with pressure transducers. For this, the Savitzky-Golay convolution filter was used, adapting the polynomial degree and convolution coefficients to reduce fluctuations <0.01 m (Savitzky & Golay, 1964). After this preprocessing step, a second set of weekly time series (W) was derived, by sampling from the daily records the groundwater head each Wednesday to simulate a weekly measurement scheme. Next, 46 groundwater dynamics indices were calculated for each hydrograph (Table 1) based on the methodology in Heudorfer et al. (2019). Indices are calculated on either daily or weekly data sets, as indicated in parentheses behind index names. Using dynamics indices instead of entire time series results in a lower-dimensional representation of the hydrographs (46×341 matrix). Indices express the entire range of groundwater hydrograph dynamics in single values. Groundwater hydrographs are viewed from three different dynamics aspects: *structure* (global hydrograph behavior), (statistical) *distribution*, and single and event-scale *shape* according to the typology by Heudorfer et al. (2019). Within each aspect, the dynamics are further divided into a number of components (Table 1) with associated indices. For example, the Richards-Baker index belongs to *flashiness* within *structure* (Table 1). As shown by Heudorfer et al. (2019), there is redundancy among the indices, but nonetheless, all indices are analyzed in order to find the indices with the strongest relations to predictors. In this paper, indices within the *structure* components are discussed more in-depth (section 4), due to their clear interpretability.

Table 1
Groundwater Dynamics Typology and Indices Used in This Study

Aspect	Component	Index
Structure	Amplitude	amp.max (D)
	Seasonality-magnitude	cvmon.min (W), parde (W), iaf.s (W)
	Seasonality-timing	colwell. M (W), vardoy.min (W), ^a vardoy.max (W) ^b
	Flashiness	BLI (D), pathlength (D), baker (D), rev.avg (D), rev.cv (D), lyapunov (W)
	Interannual variability	BLS (W), iaf.y (W), pulse.count.l (W), pulse.count.h (W), hurst (W), autocor (W)
Distribution	Density	dc.slp.m (D), avg.ann.max (D), dc.rng.01.09 (W), dc.rng.02.08 (W), dc.rng.025.075 (W), l2(W), l4 (W), colwell. C (W)
	Boundedness	dc.slp.l (D), dc.slp.u (D), l1 (W), l2 (W), median (W), peakts.avg (D)
	Modality	bimod (W), dip (W), bandwd (W)
Shape	Scale	peakbase.avg (D), sqr.avg (D), pulse.dur.l (W), pulse.dur.h (W)
	Slope	reces.const (D), recov.const (D), rise.avg (D), rise.cv (D), fall.avg (D), fall.cv (D)

Note. See Heudorfer et al. (2019) and supporting information for details.

^aIn Heudorfer et al. (2019) vardoy.min is called varjuld.min and BLI is called BFI. ^bvardoy.max is introduced in this study and is the average day of year where the hydrograph yearly maxima occurs.

2.2. Physiographic and Climatic Descriptors

Four different sets of descriptors were compiled for each of the well locations: geological, topographical, boundary, and climatic descriptors (Table 2). When considering the descriptors used, it is important to note that the aim of this initiative is not to develop predictive models at the locations where observation wells exist but for locations without observations “ ungauged aquifers ” in the sense of PUB. Therefore, the descriptors compiled do not include properties that can only be inferred from direct measurements such as aquifer transmissivity and storativity.

Local geological conditions were defined based on stratigraphy given in borehole logs. Aquifer stratigraphy and thickness (A_thickness) were interpreted in combination with information on the screen location, where available. The stratigraphy was further used to classify the aquifer state: If confining layers such as silt, clay, or marl are present and the fifth percentile of groundwater depth is above the lower boundary of a confining layer, the aquifer was judged as confined. A number of wells had no clear state, with the groundwater level sometimes above and sometimes below the confining layer and were classified as confined. Further, the cumulative thickness of the confining layers was calculated (Silt_m, Clay_m, and Marl_m). The mean thickness of the saturated zone (Aquifer_Depth) was estimated by defining the aquifer's upper bound as the mean groundwater depth (Depth_to_GW) or the bottom level of the confining layer for unconfined and confined aquifers, respectively. The aquifer's lower bound was based on low-permeability layers (silt, clay, marl, and basement rock) below the aquifer material. In a number of cases, this information was not available, when the drilling process was stopped before reaching such quasi-impermeable layers. In these cases, the lowest level in the bore log is taken as the aquifer bottom.

The morphologic characteristics slope, topographic wetness index (TWI), curvature classification, convergence index, and catchment area divided by flow accumulation were calculated from the 3-arcsec USGS HydroSheds DEM product (Lehner et al., 2008) using SAGA-GIS (Conrad et al., 2015). Subsequently, the mean and skewness of these descriptors were estimated within 300-m and 3-km buffers surrounding the wells. The two averaged areas represent conditions for recharge and groundwater flow at the local and regional scales. This heuristic approach was chosen since time series from groundwater observation wells do not capture the integral response of a well-defined catchment area as in streamflow but are point measures within an unknown aquifer volume (cf. Tiedeman et al., 1998).

Third, seven hydrologic boundary descriptors were based on topographically indicated locations of flow divides and drainage boundaries (Figure 3). Estimating these boundaries was necessary as reliable information on aquifer extents, hydraulic gradients, and boundary conditions are rarely available. The gradient to the discharge boundary was estimated by distance and height difference from the OW location/elevation to the nearest river segment (Vogt et al., 2007), as produced by the “ snap points to lines ” tool in QGIS 2.18 (QGIS Development Team, 2018). To evaluate the gradient to the recharge boundary, its location needs to be estimated first. This follows the assumptions made by de Graaf et al. (2015), who distinguished

Table 2
System Characteristics Used in the Study

Variable	Description	Range		Unit
		Minimum	Maximum	
A_thickness (G)	Aquifer thickness	0	182	m
Depth_to_GW (G)	Depth to groundwater head	0	52	m
Screen_length (G) ^b	Length of well screen	1	172	m
Screen_Upp (G)	Upper level of well screen	1	167	m
Screen_Low (G) ^b	Lower level of well screen	1	239	m
Aquifer_Depth (G)	Thickness of unsaturated zone	3	246	m
GW_Level_mean (G)	Elevation of mean groundwater level	309	932	m a.s.l.
Silt_m (G)	Cumulative thickness of silt within profile	0	8	m
Peat_m (G)	Cumulative thickness of peat within profile	0	2	m
Clay_m (G)	Cumulative thickness of clay within profile	0	14	m
Marl_m (G)	Cumulative thickness of marl within profile	0	26	m
arbel (M)	Area below observation well (OW): percentage of area below OW height	0/0	100/100	%
hgthps (M)	Position of OW relative to the height range	0/0	100/100	%
cncv (M)	Concavity: percentage of area with concave structure	0/0	100/43	%
cnvx (M)	Convexity: percentage of area with convex structure	0/0	100/44	%
cnvr (M) ^a	Mean convergence index	−11/−1 (−3/−0.5)	9/1 (3.6/0.3)	—
slp (M) ^a	Mean slope	0/1 (0/−0.5)	13/27 (3.8/2.6)	—
twi (M) ^{a,b}	Mean value of topographic wetness index	5/6 (−0.3/−0.1)	10/9 (0.5/0.3)	—
ctchr (M)	Catchment area/flow accumulation: maximum value of catchment area/flow accumulation	30,234/120,934	1,070,269/3,120,105	m ² /max number of grid cells contributing to each grid cell within radius
dist_boundary (B)	Estimated distance from the GW OW to the closest segment of the outer aquifer boundary	31	9,928	m
dist_stream (B) ^b	Estimated distance from the GW OW to the nearest stream (first- and second-order rivers; see Figure 2)	6	10,958	m
height_ow (B)	Estimated elevation of GW OW	310	950	m a.s.l.
height_stream (B)	Estimated elevation of closest stream segment	309	944	m a.s.l.
height_to_stream (B)	Meters that the stream is below/above the GW OW (gradient to stream)	−95	82	m
height_boundary (B) ^b	Estimated elevation of closest segment of outer aquifer boundary	313	967	m a.s.l.
height_to_boundary (B)	Meters that the aquifer boundary is below/above the OW (gradient to upper boundary)	−24	55	m
P_avg (C)	Mean annual precipitation	663	1,649	mm
T_avg (C)	Mean annual temperature	3.5	9.3	°C
PE_avg (C)	Mean annual potential evaporation transpiration, calculated from temperature and latitude with the Thornthwaite equation	503	623	mm
AI (C) ^b	Aridity Index	1.1	3.0	—
meanPcoldest (C) ^b	Mean precipitation in the coldest quarter of the year	41	114	mm
meanPwarmest (C) ^b	Mean precipitation in the warmest quarter of the year	72	182	mm
meanPdriest (C) ^b	Mean precipitation in the driest quarter of the year	30	80	mm
meanPwettest (C) ^b	Mean precipitation in the wettest quarter of the year	83	198	mm
ratioPcoldwarm (C) ^b	meanPcoldest divided by meanPwarmest	0.38	0.7	—
ratioPdriywett (C) ^b	meanPdriest divided by meanPwettest	0.28	0.41	—
tdiffdrywett (C)	Time difference (in months) between the driest and wettest quarters	4.3	5.8	Months
SI (C)	Seasonality index of precipitation	0.11	0.35	—

Note. Class of descriptor is indicated in brackets after the variable name. All topographic descriptors are calculated at a local scale ($r = 300$ m) and a regional scale ($r = 3,000$ m). Class of variable in parentheses: G, geology; T, topography; B, boundary, and C, climate. Data Sources: Geology: bore profiles supplied by LfU, Bavaria; topography and boundary: HydroSheds 3 arcsec; and climate: E-OBS (Europe, 0.25° grid).

^aSkewness was calculated separately from the area mean. For these, the range values are given in parentheses. ^bRemoved after prescreening for multicollinearity.

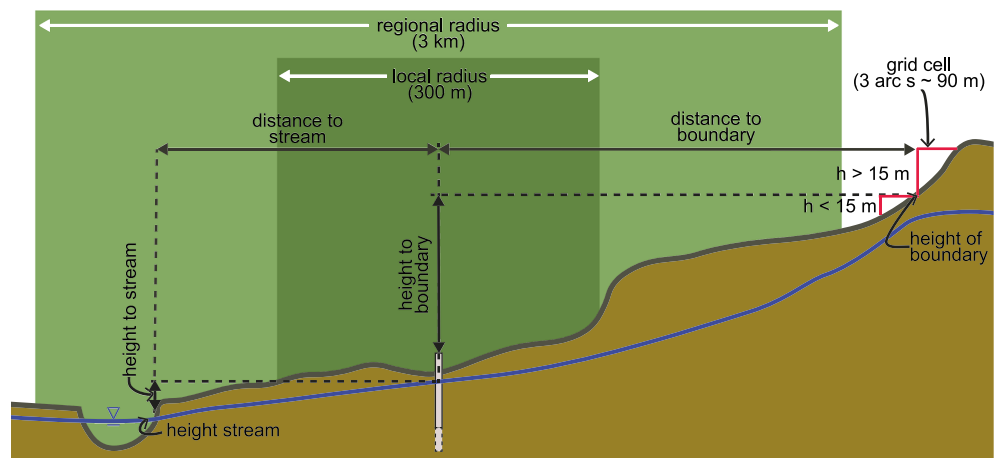


Figure 3. Principle of estimating topography-derived hydrologic boundary and morphology descriptors for each well location. Distance and height to and height of boundary are based on a recharge boundary, whereas distance height of and height to stream are based on the nearest stream segment. Shaded areas show the local and regional radii for morphology descriptors (approximate grid cell size 90 m).

sediment basins as potential aquifers from bedrock mountain ranges based on gradient changes in surface elevation. In practice, the elevation differences between neighboring grid cells in the HydroSheds DEM were calculated. When the elevation difference between two adjacent grid cells exceeds a threshold of 15 m, the “upper” boundary of the aquifer (i.e., its “recharge boundary”) was set between these two grid cells (see Figure 3). The grid resolution, however, is a sensitive parameter for finding aquifer boundaries via a threshold value (de Graaf et al., 2015). Therefore, a threshold value was found that resulted in aquifer boundaries which best matched expectations (results are shown in the supporting information Figure S1). The gradient to the recharge boundary was subsequently estimated by distance and height difference from the OW location/elevation to the boundary.

Finally, climatic descriptors were derived for each well from the 0.25° E-OBS grid (Version 17.0) covering the period 1 January 1950 to 31 December 2017 (Haylock et al., 2008). The closest grid cells of the E-OBS data set were assigned to wells based on location. Then, the reference period to calculate climatic descriptors was selected for each well, based on the selected measurement period of the well. Within this reference period, 12 climatic descriptors were calculated as described in Table 2.

2.3. Analysis of Controls on Groundwater Dynamics

In order to find controls on groundwater dynamics, first, descriptors were prescreened for collinearity. Second, correlations between all dynamics indices and descriptors were estimated for linear (Pearson correlation), monotonous (Spearman rank correlation), and nonlinear relationships (distance correlation). Cluster analysis was performed on the correlation matrices using Ward clustering with Euclidean distance. The correlation matrices and clusters are displayed in heatmaps to identify groups of similar relations.

Forward stepwise regression is often used for the identification of significant associations between a response and potential controls. Stepwise regression can be traced back to Beale et al. (1967) and is frequently used within hydrological sciences (cf. Bloomfield et al., 2009; Kuentz et al., 2017). The forward selection procedure was used to find linear associations between individual dynamics indices and physiographic and climatic descriptors. Starting from an empty model, the procedure iteratively builds a regression model from the set of descriptors. In the statistical modeling, the descriptors become predictors X to optimize the fit to a response vector Y . The response vector in this case is a single dynamics index. New predictors are added at each step from X by selecting the variable that yields the maximum absolute correlation with the residual (after orthogonalization of the variables with respect to the current model). Computation of p -values based on the t -test in stepwise regression does not account for the effects of the sequential selection of variables in stepwise regression. This leads to an overestimation of the strength of apparent relations and thus yields nonsignificant predictors (Taylor & Tibshirani, 2015). Therefore, a framework is used for adjusting the

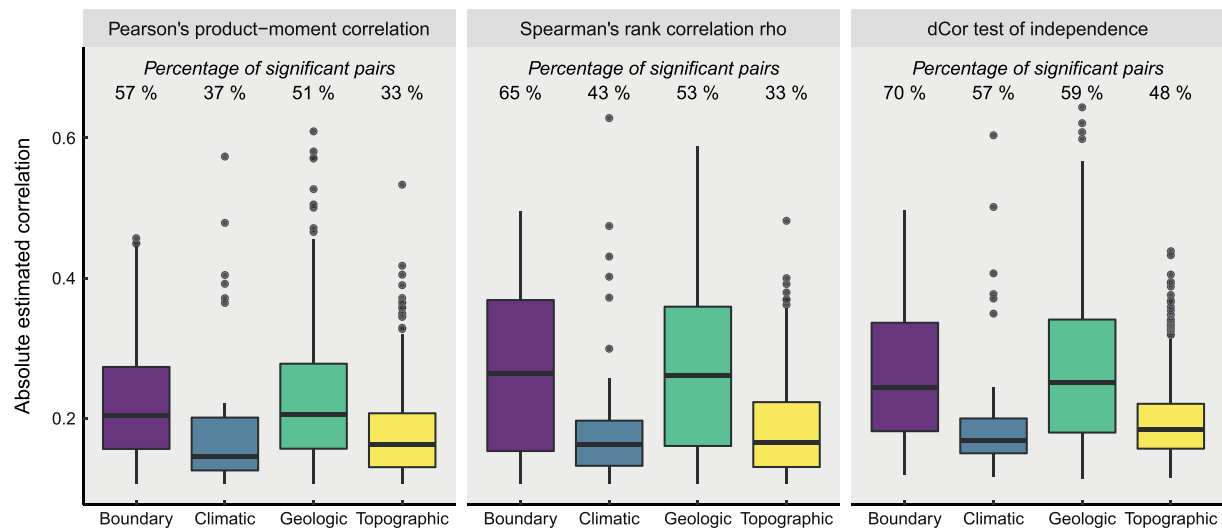


Figure 4. The boxplots show the distribution of Pearson, Spearman, and distance correlations ($p < 0.05$) between groundwater dynamics indices and descriptors associated with individual descriptor classes. Both Pearson and Spearman correlations can be negative; therefore, absolute estimated correlation is shown for better interpretability. The percentage of significant pairs shows the fraction of pairs of descriptors and indices with significant correlation.

significance of descriptors with selection-adjusted p -values computed after each step. This means that the p -value of each newly introduced predictor is computed conditionally on the coefficients of predictors entered at an earlier step as described in Tibshirani et al. (2016). Selection of predictors for each index is stopped according to a stopping rule based on the false discovery rate. The false discovery rate is the ratio of false-positives among the tests. Here a target value of ($\alpha > 0.1$) was chosen, meaning that the selection procedure is stopped, when the p -value of the latest added predictor has a 1 in 10 chance to be a false-positive (G'Sell et al., 2016; Taylor & Tibshirani, 2015). Prior to the variable selection, predictors were centered (for interpretability) and scaled (division by standard deviation) due to the predictor's different units. Stepwise regression and selective inference were performed using the R package "selectiveInference" (Tibshirani et al., 2017).

For models, where response variables can successfully be linked to controls, relative variable importance is computed by calculating standardized regression coefficients. The model selection was carried out once globally for the entire data set and then stratified by two aquifer states, confined and unconfined. Finally, the regression models for five selected indices were juxtaposed with physical process understanding in confined and unconfined aquifers. These five represent indices assigned to each of the five components of the dynamics aspect *structure* (Table 1). For each index, a conceptual model is described based on the significant predictors, signs of coefficients, and literature. The number 5 is a compromise to being able to give the analysis some depth within the scope of the paper and still show some variation across dynamics components.

3. Results

3.1. Evaluation and Redundancy Analysis of Descriptors

Multicollinearity was identified within each descriptor class (boundary, climatic, geologic, and topographic) by estimating correlation matrices and evaluating scatterplots (see sections S3–S6 of the supporting information). A total of 13 relations between the 46 individual dynamics indices and the remaining 41 descriptors (see Table 2) were studied by means of correlation analysis with Pearson, Spearman, and distance correlation analysis. The results of the correlation analysis for confined and unconfined aquifers are summarized in Figure 4; overall absolute significant ($p > 0.05$) correlations vary between 0.11 and 0.64 for all three employed methods of correlation analysis. Boundary and geologic descriptors were most strongly and frequently correlated with the groundwater indices (Figure 4). Both climatic and topographic descriptors were less strongly and less frequently correlated with the groundwater indices. Further, all correlation distributions are right skewed (except Spearman with boundary descriptors), meaning that the bulk of index-descriptor relations are relatively weak, with a few strong correlations.

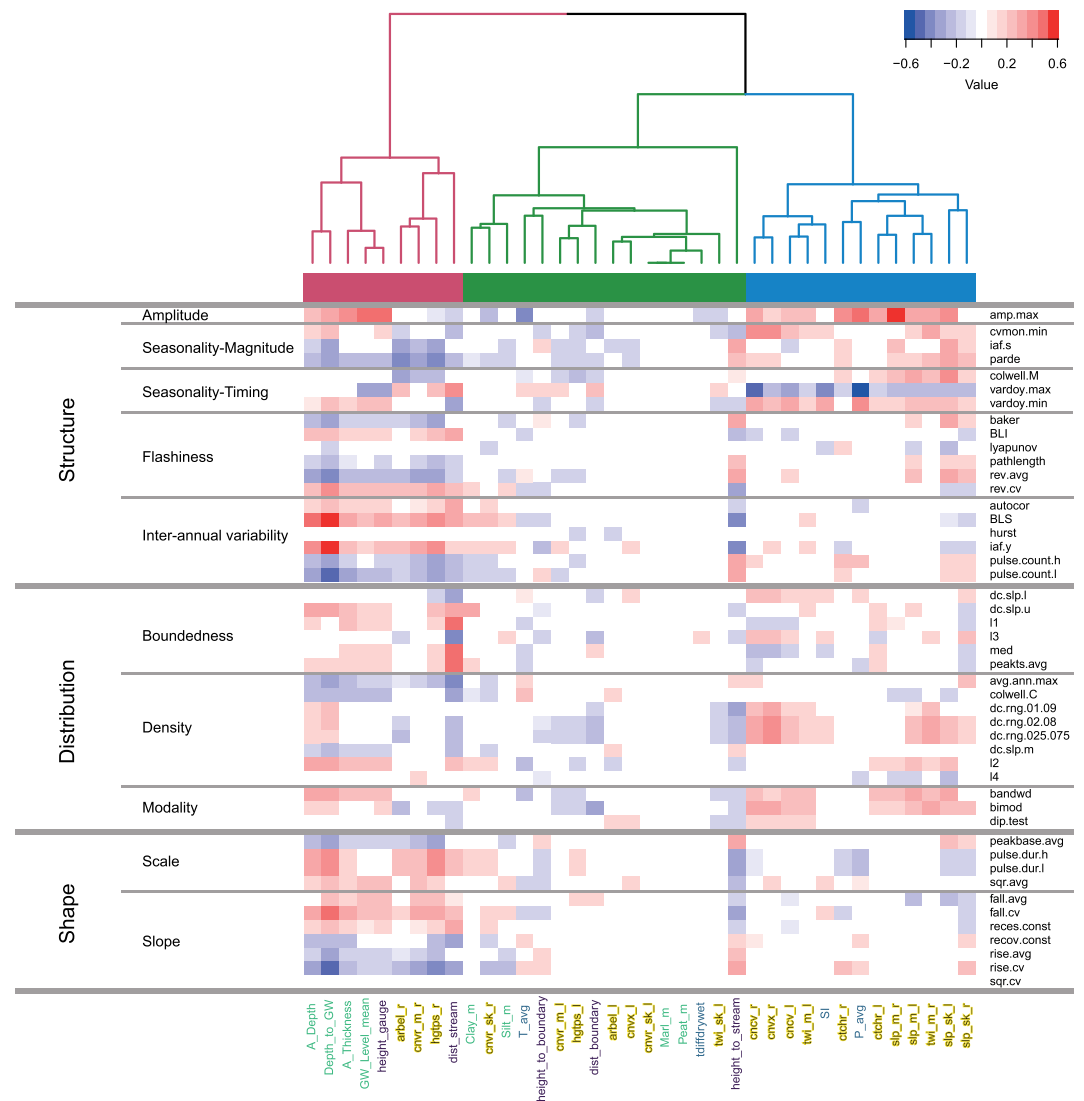


Figure 5. Heatmap of Pearson correlation (r), between groundwater dynamics indices and descriptors (details on descriptors can be found in Table 2). Blue shaded tiles describe negative correlations, and red shaded tiles describe positive correlations (white are nonsignificant relationships). Three major clusters of descriptor-index relationships can be discerned, as indicated by the dendrogram and colored bands (red, green, and blue) looking at the columns.

The heatmap of Pearson correlations shows all individual relations (Figure 5). The red cluster with the strongest relations comprises mainly geologic descriptors and also a few descriptors from boundary and topographic classes. The green cluster contains descriptors from all four classes but shows few significant and mostly weak relations with the exception of height to stream. The third cluster (blue) comprises mainly topographic descriptors and also precipitation and seasonality of precipitation. Most indices have weak relationships to topographic descriptors. Reading row-wise, certain similarities within dynamics components can be seen, although sometimes with inverse correlation, for example, *flashiness*, where correlations between indices and descriptors are similarly strong, except for the Lyapunov exponent. The same goes for *interannual variability*, where similarity is strong, except for the hurst coefficient. Other components have more diverse patterns, such as *seasonality-timing* or *density*, meaning that indices within this class are correlated to different characteristics.

Spearman and distance correlations were similar, although for both there was a larger number of significant pairs and higher absolute correlation coefficients. To put this in numbers, Pearson correlation coefficients were estimated between the Pearson correlation matrix and both distance and Spearman correlation

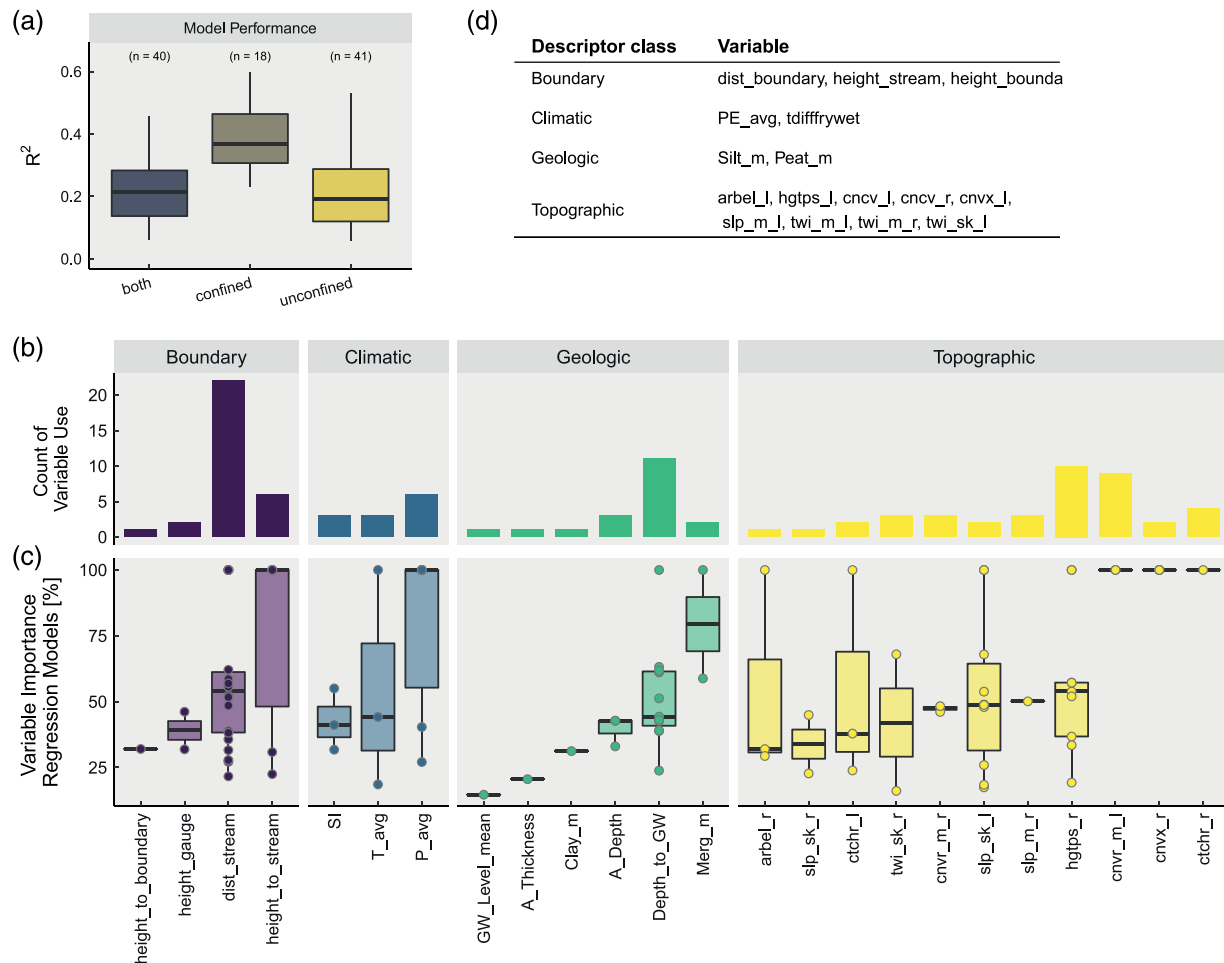


Figure 6. (a) Model performance (coefficient of determination, R^2) of all regression models, where statistically significant relations between descriptors and individual groundwater dynamics indices could be found. The coefficient of determination is shown differentiated by wells in confined and unconfined aquifers. Above each boxplot, the number of models within is given. (b) Absolute count and (c) relative importance (based on the slope of the regression coefficients) of predictors in exploratory regression models across all models. (d) Nonsignificant descriptors from model selection for confined and unconfined aquifers.

matrices. The similarity among the three methods is high ($r = 0.8$), with the exception of a few descriptor-index relations that show large increases in correlation when using Spearman or distance correlation. Among these, distance and height to stream, depth to groundwater, and skewness of slope on the regional scale show increased correlations (more than 0.2 higher) for distance correlation. For Spearman correlation, the same descriptors and thickness of saturated zone, height position of OW, and median slope on the regional scale have comparably higher correlation coefficients (see supporting information). In summary, when comparing linear with nonlinear/monotonous dependencies, many important descriptors (red and partially blue clusters) also show significant nonlinear behavior.

3.2. Regression Analysis on Dynamics Indices and Predictors

Forward stepwise regressions were built for each index individually using a combination of descriptors. The selection procedure successfully identified regression models with significant predictors for all three groups: all wells, confined aquifers, and unconfined aquifers (Figure 6a). Median model performance for data from confined and unconfined aquifers is relatively low with a median R^2 of 0.21 for all wells and a median R^2 of 0.20 for unconfined-only wells. Regression models for 40 out of 46 indices were significant with at least one predictor (compare Table S2 of the supporting information). Models identified the confined aquifers performed better with an average R^2 of 0.42, although fewer models (18) could be found. The number of significant predictors across all models varies between 1 and 5, although for confined aquifers, models never had

more than two predictors (supporting information Figure S8). Model performance improved with the number of predictors and models with only one predictor generally having low R^2 , which is not unexpected given relatively low correlation coefficients between indices and many individual descriptors as shown in section 3.1.

Figure 6b shows how frequent predictors were selected for all unconfined and confined models that have at least one significant predictor. It can be seen that 24 (out of a total of 38) predictors were statistically significant in at least one model, while 18 were significant in multiple models. Among boundaries, the most frequently selected predictor was *distance to stream*, which was also the overall most important predictor of groundwater dynamics. *Distance to stream* was used in 23 models and has a median importance of 50% in these models (Figure 6c). *Height difference to stream* was significant in five models and had in more than half of the cases 100% importance, meaning it was the sole variable in these model. Among climate characteristics, average temperature, precipitation, and the seasonality index were significant two, five, and three times, respectively, and generally were among the less important variables (median below 50% importance). Within geology, *depth to groundwater head* was selected most frequently (in 11 models) and was the second most important predictor overall. *Aquifer depth* was the second most important predictor within geology but was not that important in general. Among topography, the most frequent predictors were *mean of regional slope* (used in nine models) and regional catchment area divided by flow accumulation (used in 10 models) (Figure 6c). When investigating potential predictors that were not selected in any model, the topographic class had the highest count. Here only roughly half are significant (11 out of 21), where especially predictors on the local scale were nonsignificant (Figure 6d). Among the boundaries, distance and height to boundary were nonsignificant, while height difference between OW and boundary was used in one model only.

4. Process-Based Assessment of Selected Index-Control Relationships

In this section, a subset of indices and the respective models built with selective inference are discussed. A selection was made to describe different hydrograph aspects (amplitude, magnitude, timing, flashiness, and interannual forcing; see Table 1). Further, among the indices associated with these dynamics, the index was selected from a model with relatively high performance (above median) as well as from viable models for confined and unconfined aquifers. Table 3 shows the selected models and their performance with mean $R^2 = 0.46$ and $R^2 = 0.37$ for confined and unconfined aquifers, respectively. For all indices except interannual fluctuation (iaf.s), models for the sample of confined and unconfined aquifer records have higher R^2 than have models fitted to all records. This confirms the expectation that confined and unconfined aquifers are affected by different site characteristics. Table 3 also shows that 5 predictors are used for confined aquifers and 11 predictors are used for unconfined aquifers. Among the five selected indices, *distance to stream* (dist_stream) and *average annual precipitation* (P_avg) are again the most frequent predictors, followed by *height position of OW* (hgtps_r) and *median convergence regionally* (cnvr_m_r). From Table 3 it can be seen that in accordance with the fact that confined aquifers are not recharged locally, only regional variables are significant for confined aquifers. *Local convergence* (cnvr_m_l) is an exception and is discussed below.

The next five sections describe the different models for the subset of indices by relating significant predictors to literature and conceptual thinking to index-specific conceptual models. Additionally, the process-based assessment is supported by connecting the conceptual models to the spatial patterns of the indices. To facilitate the reader's understanding, the predictors are shown in illustrations within a conceptual landscape in the center for unconfined and to the right for confined aquifers (Figure 7). Additionally for each of the five indices, time series that correspond to minima and maxima (end-members) of index values within the data set are plotted. These show the extremes of groundwater dynamics for each index (Figure 7, left column).

4.1. Amplitude: Groundwater Level Amplitude (amp.max)

Groundwater level amplitude is defined as the difference in amplitude between the lowest and highest measured groundwater levels. Figure 7a-I shows the hydrographs with the highest (blue) and lowest (green) amp.max within the data set for unconfined aquifers and confined aquifers. The figure shows that amplitudes are higher in unconfined aquifers, as seen by the upper end-members. Mathematically, the amplitude of unconfined aquifers can be described (in order of decreasing importance) with the predictors *mean slope on the regional scale* (slp_m_r), *average precipitation* (P_avg), and *skewness of slope at the regional scale* (slp_sk_r) (the standardized coefficient value can be interpreted as importance of respective predictor).

Table 3
Parameters and Coefficients of Selected Multiple Regression Models Identified by Selective Inference

Index	Aquifer	Type	Category	R^2	Term	Coefficient	Significance
amp.max	Confined	Structure	A*	0.49	(Intercept)	1.21	****
					P_avg	0.67	***
	Unconfined	Structure	A*	0.45	(Intercept)	2.47	****
					slp_m_r	1.19	****
					P_avg	0.65	*
iaf.s	Confined			0.45	slp_sk_r	−0.49	*
					(Intercept)	0.46	****
					hgtps_r	−0.06	*
	Unconfined			0.21	cnvr_m_r	−0.06	*
					(Intercept)	0.50	****
vardoy.max	Confined			0.46	hgtps_r	−0.04	*
					slp_sk_l	0.04	*
					(Intercept)	79.98	****
	Unconfined			0.47	P_avg	−13.58	**
					(Intercept)	63.24	****
BLI	Confined	Structure	Flashiness	0.30	P_avg	−11.06	****
					SI	−8.25	**
					dist_stream	7.48	**
	Unconfined	Structure	Flashiness	0.20	(Intercept)	0.96	****
					cnvr_m_l	0.02	*
BLS	Confined	Structure	IAF*	0.54	(Intercept)	0.95	****
					dist_stream	0.02	**
					height_to_stream	−0.01	*
	Unconfined	Structure	IAF*	0.53	A_Thickness	0.01	*
					(Intercept)	0.43	****
	Confined			0.54	dist_stream	0.11	*
					cnvr_m_r	0.08	*
					cnvx_r	0.06	****
	Unconfined			0.53	(Intercept)	0.30	****
					Depth_to_GW	0.08	***
					dist_stream	0.05	*
					hgtps_r	0.04	*
					GW_Level_mean	0.03	****

Note. Predictors for confined aquifers: P_avg (two times), hgtps_r, cnvr_m_r (two times), cnvr_m_l, dist_stream, cnvx_r. Predictors for unconfined aquifers: P_avg (two times), SI, slp_m_r, slp_sk_r, slp_sk_l, hgtps_r (two times), dist_stream (three times), height_to_stream, A_Thickness, Depth_to_GW, GW_Level_mean.

* $p \leq 0.05$.

** $p \leq 0.01$.

*** $p \leq 0.001$.

**** $p \leq 0.0001$.

The coefficients slp_m_r and P_avg are positive, meaning that their increase leads to an increase of amplitude (Table 3). This means that high amplitudes are reached in areas with steep surrounding slopes and high amounts of annual precipitation. The third most important predictor *slope skewness at the regional scale* is negatively correlated to amplitude (slp_sk_r can be both positive and negative, due to the definition of skewness). High skewness in slopes indicates abrupt topographic changes, such as at the edge of the valley, with high slopes (coincides with the importance of slp_m_r), followed by relatively flat areas. This type of geomorphological setup is illustrated in Figure 7a-II and typical of mountain valleys. Examples of this situation are areas where mountain block recharge and front slope flow occurs (Markovich et al., 2019; Wilson & Guan, 2004). The higher amplitudes can also be explained by the findings by Gleeson and Manning (2008), who showed that groundwater flow in areas of increased valley incision is predominantly local (front slope flow). This groundwater recharge into such narrow mountain valleys cannot dissipate laterally as in more extensive regional aquifers, consequently causing larger amplitudes. Looking at the other extreme, low amplitudes occur in systems with a rather flat regional topography and low precipitation. In these basin-fill groundwater systems, intermediate and regional flow systems are important (Tóth, 1963). This is corroborated by the spatial pattern of amp.max (Figure 8a).

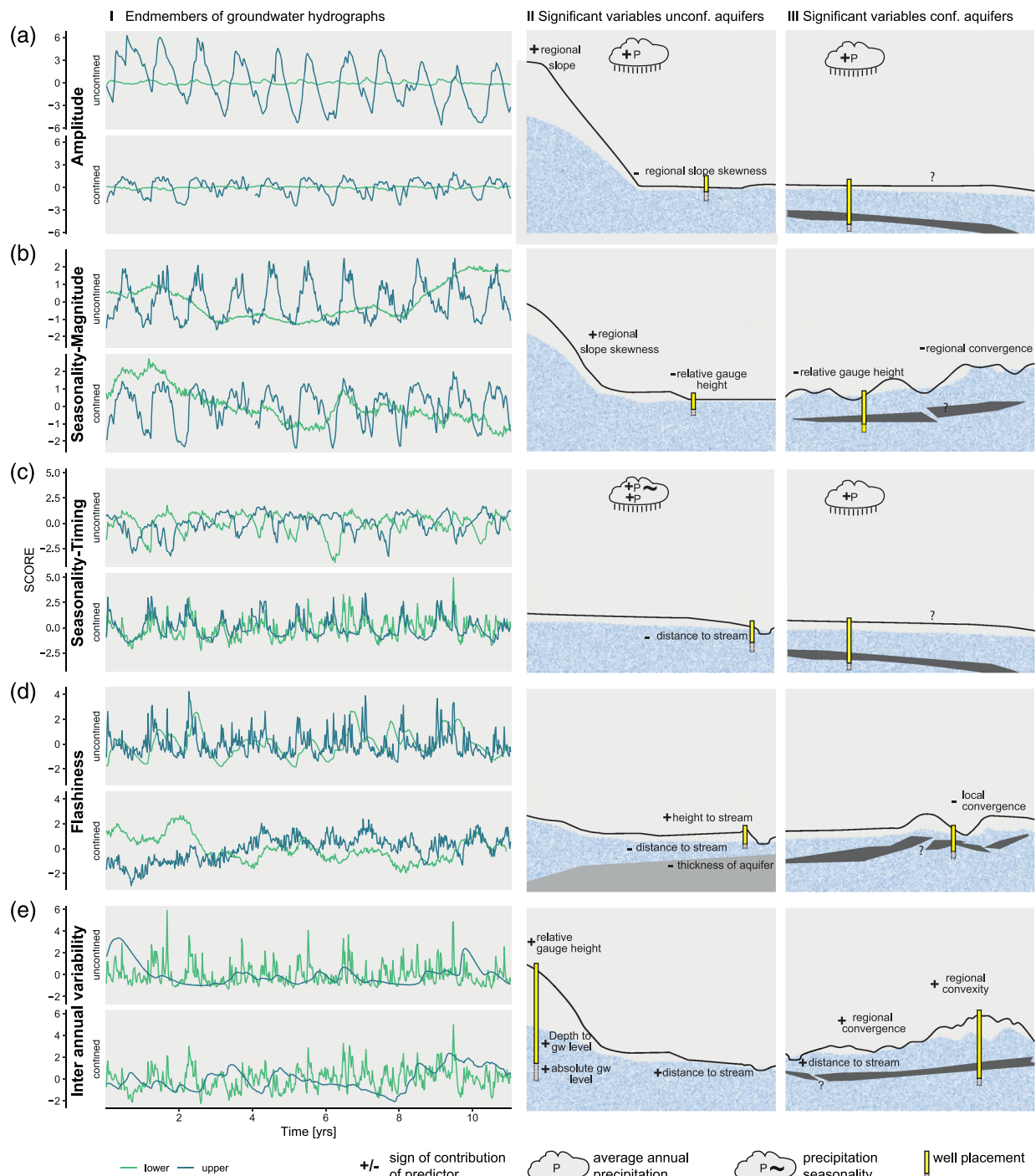


Figure 7. Indices and conceptual model explanations. Well placement indicates the position in the landscape where high values of the individual dynamics component is expected. All hydrographs in column I are centered and scaled by standard deviation except for the panel showing amplitude.

The highest values are found at the foot of the Alps in the south and the low values in the alluvial river valleys and outwash plains.

In confined aquifers, none of the geologic or topographic predictors play a role and only *average annual precipitation* is significantly correlated with groundwater level amplitude (Figure 7a-II). The recharge pathway does not follow from the regression model as indicated by a question mark in Figure 7a-III. Recharge of confined aquifers is usually thought to occur by seepage from overlying layers or in recharge areas at higher

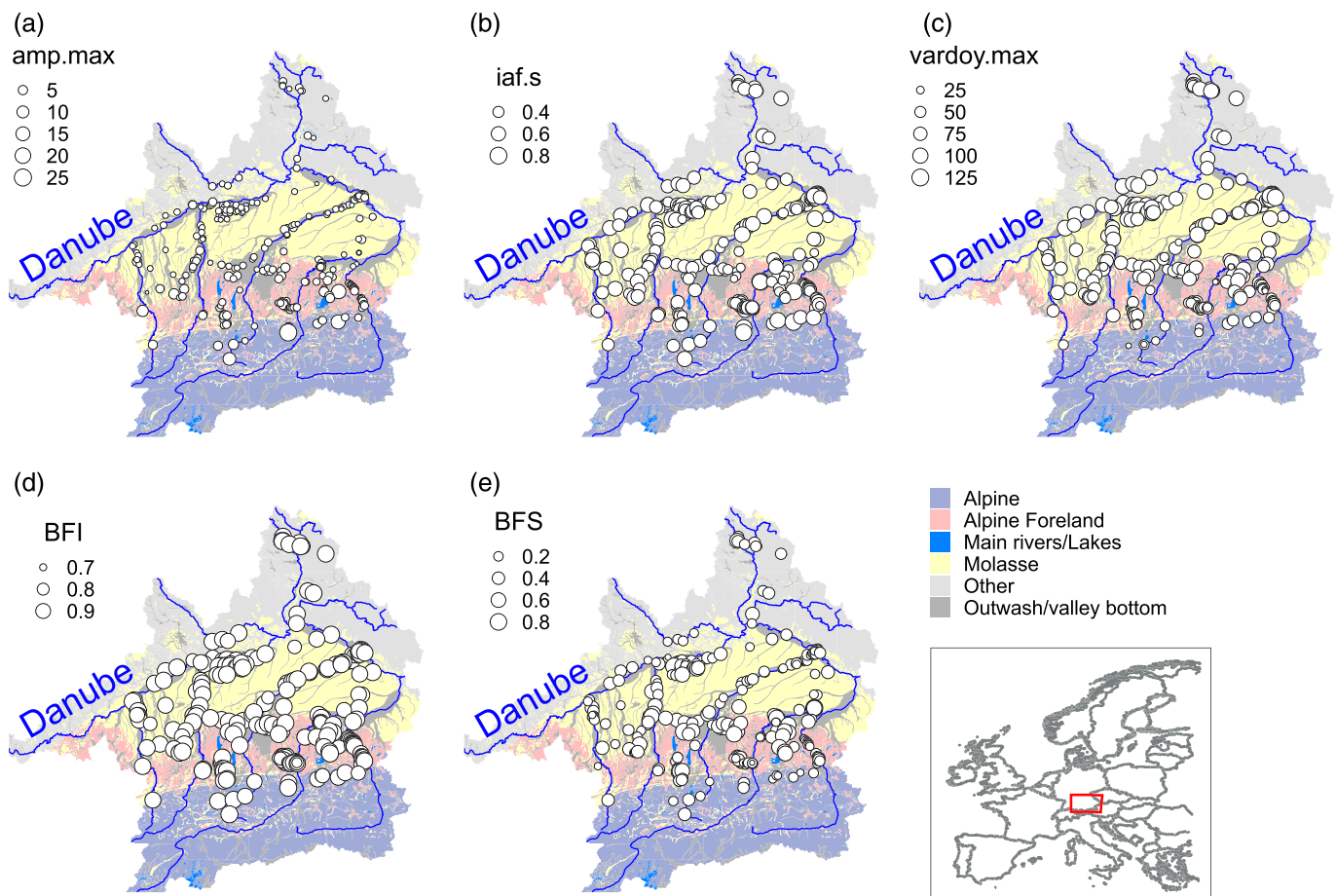


Figure 8. Spatial pattern of groundwater dynamics in unconfined aquifers as summarized with five selected indices representing (a) groundwater amplitude, amp.max; (b) seasonality-magnitude, iaf.s; (c) seasonality-timing, vardoy.max; (d) flashiness, BFI; and (e) interannual variability, BFS.

elevations, where the aquifer crops out (Fetter, 1994). Here annual average precipitation values are assigned to each well and are therefore relatively local (20-km grid cell). On the other hand, mean annual precipitation is spatially smooth and should be representative of input signal to confined aquifers.

4.2. Seasonality-Magnitude: Average Seasonal Fluctuation (iaf.s)

Average seasonal fluctuation expresses the regularity of the variation in water levels from year to year. The variability of the seasonal magnitude can be seen in Figure 7b-I, where the seasonal magnitude of the amplitude of the upper end-member (blue) for both confined and unconfined aquifers is very regular. The lower end-members (green) show no clear average seasonal fluctuation but a higher inter-annual fluctuation. According to the regression results, average seasonal fluctuations (iaf.s) can be described for unconfined aquifers by *relative height of observation well* (hgtps_r) and *regional slope skewness* (slp_sk_r), as shown in Table 3. High average seasonal fluctuation is thus expected for locations situated in a depression and locally surrounded by topographic ridges. This can be attributed to wells located in floodplains of incised smaller river channels (not captured by first- and second-order rivers [classic stream order] used for the predictor distance_to_stream), where the groundwater is heavily dominated by streamflow as illustrated by Figure 7b-II. In the study area, floodplains and rivers are continuously fed by lateral groundwater flow from the upland during late spring and summer. In autumn and winter, yearly floods occur; e.g., snowmelt-driven floods approximately at the same time each year are likely the driver of high regularity of seasonal fluctuation in groundwater through fast propagation of the flood signal into the groundwater and surface water-groundwater interaction. Propagation may occur through, e.g., floodwave propagation in the gravelly parent material (Buffin-Bélanger et al., 2015) or groundwater ridging (Sklash & Farvolden, 1979), where groundwater levels reach the loess and silty soils, which are predominant in the riverine valleys

(BGR, 2018). Spatially, the data set shows a very heterogeneous pattern of average seasonal fluctuation (Figure 8b). High values can be seen both along the river valleys and in the pre-Alps in the south.

Figure 7b-III shows that for confined systems the best model describing average seasonal fluctuations uses the predictors *regional convergence* (*convr_m_r*) and *relative height of observation well* (*hgtps_r*). Both coefficients are negative, meaning that low relative OW elevation in a regionally undulating terrain (negative regional convergence) can be linked to high regularity of seasonal fluctuation (Figure 7b-III). Any clear explanation of this relationship could not be found, but an undulating landscape (*convr_m_r*) may indicate connectivity between the overburden and the confined aquifer, enabling local recharge to the confined aquifer in depressions. The result of this is similar to unconfined aquifers, extremes of water availability being more pronounced at low well elevations within the landscape. Spatial patterns were not seen in the data set.

4.3. Seasonality-Timing: Variance in Day of Year of Annual Maxima (*vardoy.max*)

The regularity of seasonal timing within the year is quantified by *vardoy.max*. It expresses the variance in day of year when the annual maximum groundwater head occurs. High values of *vardoy.max* express low regularity in seasonal timing. Figure 7c-I shows the difference of seasonality-timing between time series with high (blue) and low (green) regularity of timing. The example of an upper end-member (blue) for unconfined aquifers is relatively irregular with regard to magnitude, but timing of the maximum levels occurs approximately at the same time each year. Looking at the regression model for unconfined aquifers, *average annual precipitation* (*P_avg*), *precipitation seasonality* (*SI*), and *distance to stream* (*dist_stream*) were selected as predictors in decreasing order of importance (Table 3). This means that high precipitation combined with a high precipitation seasonality and a location close to streams leads to high regularity in timing of annual maxima of groundwater levels (the index *vardoy.max* has high values when seasonality-timing is low; thus, in Figure 7c-II, when predictors have low values, they are important for seasonality-timing). For unconfined aquifers on a regional scale, precipitation amount and seasonality are directly related to the amount and timing of precipitation reaching the groundwater. Further, vicinity to stream increases regularity of timing, as the lateral recharge from upland aquifers converges at the river (Heudorfer & Stahl, 2017). Further, interaction between surface water and groundwater as described above in section 4.2 may occur during yearly snowmelt. Within the study domain, the Bavarian alpine areas receive the highest amounts of precipitation, which accumulates as snow for extended periods of time. Snowmelt dominates not only high flows of river discharge but also the recharge of aquifers and occurs relatively consistently in spring, giving high regularity of timing in particular in first- and second-order streams. In the drier and warmer north, recharge is generally much lower (BGR, 2019) and less regular. This effect can also be seen in the pattern of index values in Figure 8c, where a clear northeastern-southwestern gradient emerges, with high regularity in the Alps and pre-Alps and lower regularity in the flatter parts of northern Bavaria.

For confined aquifers, timing is related to average annual precipitation (*P_avg*) as seen in Figure 7c-III, where higher amounts lead to a more regular seasonality. No clear explanation is apparent, either conceptually in literature or when studying the spatial pattern (although to some degree it follows the northeast-southwest gradient for unconfined aquifers) (Figure 8d).

4.4. Flashiness: Base Level Index

The base level index (BLI) describes the rapidness of change in groundwater levels with regard to a baseline groundwater level (Heudorfer et al., 2019). Flashiness of groundwater levels in unconfined aquifers is thus expressed by the BLI that assigns low values to flashy hydrographs. In Figure 7d-I) flashy hydrographs (blue) are contrasted with hydrographs with delayed response (green) for confined and unconfined aquifers. Flashiness is generally higher among OW in unconfined aquifers within the data set. Table 3 shows that in unconfined aquifers, high flashiness is linked to decreasing *distance to stream* (*dist_stream*) and *thickness of aquifer* (*A_depth*) and decreasing *height to stream* (*height_to_stream*) (coefficients in Table 3 and predictors in Figures 7d-II and 7d-III show reverse signs, as the BLI is large when flashiness is low). These three predictors can partially explain flashiness, using observations made related to storm events found in the literature. Focused recharge is a concept that describes the immediate rise of the groundwater table in the vicinity of rivers, following a storm event as described by Winter (2001). Further, the magnitude of the mound is inversely related to the thickness of the saturated zone (Sklash & Farvolden, 1979), here expressed in the predictor *aquifer thickness*. In accordance with the regression model Seibert et al. (2003) demonstrated on the

hillslope scale that groundwater dynamics close to the river are linked to the generally flashier streamflow dynamics. A predictor that is more difficult to merge with the aforementioned theory and observations is the contribution of a large gradient between groundwater well and river to flashiness (*height_to_stream*). It is not completely clear if this relates to deeply incised river valleys, where flashiness in groundwater levels accordingly would be higher. Deeply incised rivers imply a stronger underflow component and thus limited connectivity between river and groundwater (Larkin & Sharp, 1992). On the other hand, high flashiness without connectivity to rivers could in such cases be explained by local recharge events, an absence of stabilizing regional flow, and a short distance to groundwater table and gradients (Gleeson & Manning, 2008) on steeper slopes. Further, Voltz et al. (2013) showed that the highest groundwater table variation occurs in steep valleys at high points in the valley sides, which runs counter to the groundwater ridging model by Sklash and Farvolden (1979). The theories above partially conflict due to different scales and physical environments. Although the concepts support the regression model to some extent, they are hard to unite on the regional scale. Since the model has a relatively low R^2 and the pattern in Figure 8d does not reveal any clear explanation, high flashiness in groundwater levels is surely a consequence of multiple different factors, with explanations to be found locally.

In confined aquifers, high flashiness is coupled to high local convergence (Figure 7d-II). Wells with flashy hydrographs are thus placed in topographic depressions. Topographic convergence can indicate thrusts or terraces, where connectivity between the unconfined and confined aquifers can lead to fast recharge to the unconfined aquifer.

4.5. Interannual Variability: Base Level Stability

Inter-annual variability is represented by base level stability (BLS). BLS is calculated by taking the difference between the annual maximum and minimum BLIs within the measurement period. BLS measures the dominance of interannual variability—a large BLS results in a large dominance of interannual variation as seen in the upper end-members of time series from confined and unconfined aquifers in Figure 7e-I). For unconfined aquifers, the most important factor to model BLS is *depth to groundwater* (*Depth_to_GW*) followed by *distance to stream* (*dist_stream*), *height position within the surrounding region* (*hgpts_r_m*), and *mean groundwater level* (*GW_Level*). In Figure 7e-II) these predictors are shown together with a OW that indicates the location where high interannual variability is expected. Intuitively, deep aquifers (*Depth_to_GW*) respond to long-term changes, i.e., climate variability, due to more regional and thus slow recharge as well as longer travel time in the unsaturated and subsequently saturated zones (Haitjema & Mitchell-Bruker, 2005; Tóth, 1970). Further, it has been shown that with increasing depth of water table, a more pronounced asymmetry in the water table response to wet and dry periods is seen, which manifests in a more pronounced interannual variability (Eltahir & Yeh, 1999; Vivoni et al., 2008). When depth to water table is small and distance to drainage high (*dist_stream*), the importance of baseflow in the stream decreases (Marani et al., 2001). Thus, the closer wells are placed to streams, the more the groundwater head is influenced by aquifer interaction with the streams dynamics, which have a more pronounced yearly seasonality (Cuthbert, 2014; Dettinger & Diaz, 2000). According to the regression model, interannual variations increase with distance to stream and with relative and absolute elevations of the wells. Comparing Figure 8e with Figure 2, wells with high interannual variability are found in the outwash plains, in mountain valleys, or in pre-Alps in the upland. A possible explanation is that in mountainous terrain, multiple smaller catchments discharge into the larger alluvial aquifers of the mountain valleys, thus leading to a more stable signal over time in the valleys.

Interannual variability, expressed by BLS in confined aquifers, increases with distance to stream (*dist_stream*) and regional convergence (*cnvr_m_r*) and convexity (*cnvx_m_r*) as seen in Table 3 and illustrated in Figure 7e-III). Interpretation of these factors for confined aquifers is not straightforward, but the largest interannual variability can be found in the topmost molasse formation, when confined by old drift in areas that are covered by the rolling hills (high convexity and convergence) of young moranic landscapes.

5. Overarching Discussion

This study shows that simple empirical models can link climatic and physiographic properties to groundwater dynamics through multiple regression and that these empirical models have a physical meaning.

Further, despite a heterogeneous landscape and a large spatial scale, clear relationships can be found. Without optimization of models, groundwater dynamics components could be modeled with an average R^2 of 0.34 for unconfined aquifers and of 0.47 for confined aquifers. It can be argued that this study is a starting point for statistical regionalization of groundwater dynamics on the regional scale and thus a promising concept to develop predictive models of groundwater dynamics in locations/regions where data and information to parameterize and calibrate numerical models of groundwater flow are not available. Before regionalization can be achieved, several challenges that were encountered when designing and performing this study must be tackled.

The main challenges to this approach as means of prediction lie within a complete and optimal selection and design of physiographic and climatic descriptors. Related to this, a major aim of this study was to find relevant controls of groundwater dynamics. The selection and design of descriptors used in this study were a compromise of (typically) available data at the regional scale and quantifying the relevant factors influencing groundwater dynamics. This was partially successful, as some of the most relevant predictors in the models are clearly coupled to groundwater flow processes: distance to stream as a drainage boundary, average annual precipitation as an input signal, depth to groundwater level, indicating time lag of recharge and topographic features, such as slope characteristics of the surrounding environment, which is coupled to groundwater recharge, surface runoff and the hydraulic gradient. This shows, on the one hand, that the entire range of predictor classes (boundary, climatic, geologic, and topographic) are needed. On the other hand, this study shows that only a relatively small number of 23 out of 54 tested predictors are significant. And further, among this reduced set of predictors, only 18 are significant in more than one model. However, some uncertainties regarding completeness and optimality of these predictors remain.

Among these predictors with low significance, many belong to the class of topographic predictors. The importance of topographic controls, such as TWI, curvature, and slope, on groundwater response has been previously shown in a number of studies at the catchment and hillslope scale (e.g., Rinderer et al., 2016, 2017, 2019; Seibert et al., 2003). The present study shows only slope as an important predictor, while TWI and curvature (concavity and convexity) are rarely significant. This can be due to a reduced importance of the latter factors for alluvial aquifers on the regional scale, where a larger range of controlling factors are at play as opposed to the small-scale variability of, e.g., TWI, which controls groundwater dynamics at the hillslope scale. Another reason can be the choice of representative area, which was taken using averaged information within two different radii, one on the local scale and another on the regional scale. This was a compromise due the unknown aquifer extents. Published literature gives very little guidance at which scale morphologic characteristics should be derived. Nevertheless, the adapted approach appears successful, as some predictors such as slope and convergence are important at the regional scale (Figure 6b). In conclusion, while some predictor's derivation is straightforward, others such as within the topographic class retain some uncertainty and results cannot be linked to the published literature. Here some work is left to do, for example, a specific sensitivity study for choosing radii for the local and regional scales.

Other factors were expected to be highly relevant to alluvial aquifers but proved to be nonsignificant, such as the group of drainage divide boundaries (height_boundary, height_to_boundary, and distance_to_boundary). As shown by, e.g., Cuthbert (2014), distance to boundary is an important control of groundwater recession behavior for unconfined aquifers. As mentioned above in section 2.2, groundwater basins are hard to delineate, and a topographic proxy was derived instead. The low significance of these derived boundary predictors may indicate that the definition of boundary predictors needs improvement. Even if the design of some predictors is not straightforward, relationships between morphologic features and groundwater flow in unconfined aquifers have a more solid basis in theory (cf. Haitjema & Mitchell-Bruker, 2005; Tóth, 1963). Confined aquifers, on the other hand, are comparatively understudied, which made description of models for confined aquifers generally more difficult. Models for confined aquifers contained two significant predictors at most, due to a lack of easily derivable, relevant descriptors. Topographic predictors could only be used as proxy information for characteristics of the hydrostratigraphy, while relevant factors related to recharge mechanisms are expected underground as in confining and leaky layers. Information from borehole data does not alleviate the situation, since only the condition at the location of the observation well, not the larger structure, can be considered. Still, models for confined aquifers performed better than those for unconfined ones, despite the lack of derivable potential predictors and also in process knowledge. This indicates that the dynamics of unconfined aquifers might be easier to describe with empirical models.

Using a regression-based regionalization methodology on the regional scale is based on the assumption that any point in space can be found by continuous interpolation. This will work in cases where aquifer conditions are homogenous (cf. Cuthbert, 2014), but generally, groundwater observations are considered point observations due to the presence of heterogeneity (de Marsily et al., 2005) and consequently groundwater dynamics can be discontinuous in space (Barthel, 2014; Heudorfer et al., 2019). In other words, local heterogeneity can surely not be handled by the proposed macroscale models with only a handful of predictors. Further, models are constructed across multiple discontinuous and discontinuous aquifers with varying (and here unknown) hydrogeological properties (e.g., storage coefficients and hydraulic conductivity), which leads to different behaviors among aquifers. Apart from the abovementioned uncertainties in design of descriptors, variability of hydrogeological properties likely explain a considerable fraction of the unexplained variance and thus the moderate explanatory power of the empirical models. Because of the methodology leading to viable results despite the (common) lack of descriptors, it has strong potential to be adapted (and improved) in other similar regions. Initially, adaptation should be tested in similar humid climate and covering the same hydrological landscape types, mountain valley, riverine valley, and hummocky terrain (Winter, 2001).

Finally, it should be kept in mind that the goal of this study was not to develop the best possible, universal predictive model, but as a starting point to link a dynamics and controls at ungauged sites. Despite the strengths and limitations of the presented approach, open questions of potentially relevant missing predictors, representative scale, and nonlinear predictor-dynamics relationship should be addressed before operationally attempting prediction in ungauged aquifers. Populating predictive models based on observational studies may be aided by machine learning tools, which can result in models with higher explanatory power yet more difficult interpretability (Schmidt et al., 2020). Future work should also adopt a more advanced preclassification scheme, except separating confined from unconfined aquifers. Heudorfer et al. (2019), for example, separated three classes based on expected response: shallow systems from deep confined and deep unconfined aquifers. The example showed that linear function between indices and system descriptors can have strongly deviating slopes among the three different response classes. Other factors that may be useful for preclassification are which flow system the well is tapping as well as local similarity, whose importance on groundwater dynamics were demonstrated by Giese et al. (2020).

6. Conclusions

This study systematically investigated the statistical relation between groundwater dynamics and easily available and accessible climatic and physiographic properties across a large spatial and geomorphologic domain in southwestern Germany. Despite a heterogeneous landscape and a large spatial scale, clear relationships between site properties and groundwater dynamics emerge and have been quantified using a relatively simple methodology. With regard to research questions (1) and (2), it was shown that two sets of physiographic and climatic descriptors are more strongly correlated with groundwater dynamics than others and are among the most important predictors in regression models. Specifically, across all aspects of dynamics, hydrogeological boundaries and hydrostratigraphic (geologic) properties are the primary controls on groundwater dynamics. Among the boundary controls, groundwater well distance to stream and gradient to stream are important, while distance to groundwater and thickness of the unsaturated zone are the most important hydrostratigraphic controls. Other important controls are annual average precipitation, steepness and distribution of surrounding slopes, and flow accumulation on the regional scale, among climatic and morphologic descriptors. These results were demonstrated on a large regional data set showing strong relationships that can be used to predict groundwater dynamics at ungauged sites. To the knowledge of the authors, for the first time in the scientific literature, the strength and relative importance for different components for prediction at “ungauged” groundwater sites has been quantified.

With regard to research questions (2) and (3), it was shown that groundwater dynamics in unconfined aquifers of amplitude, magnitude of seasonality, and interannual variation can be described reasonably well with global multiple regression models within the study domain and can be based on process understanding ($R^2 = 0.5$). Certain dynamics components like flashiness and timing of seasonality may be more dependent of local conditions and are thus harder to describe ($R^2 = 0.2$). Finally, groundwater dynamics in confined aquifers, modeled with multiple linear regression models, have relatively good model fits ($R^2 = 0.3$ – 0.6).

However, it is more difficult to give physical explanations that could support the validity of those models than it is for unconfined aquifers.

In general, the literature reveals relatively little about empirical relations between groundwater dynamics and environmental controls, which are studied only for special cases and small catchment scales. More commonly, controls on local and regional recharge is studied instead, which can be used to deduce the impact on groundwater dynamics. Finally, this study shows a way forward to establish statistical regionalization of groundwater dynamics. Transfer to other regions with similar hydrogeological and meteorological conditions should be possible, since the presented models can be constructed based on a relatively small number of predictors derived from generally available borehole data, gridded climate data, and Digital Elevation Models (DEMs). More advanced predictive models can surely be established, but first after further investigating the impact of the choice of scale; missing predictors, taking into account nonlinearity of predictor-dynamics relationships; and pre-classification of sites. While tedious regarding data preparation and hence rarely done, the results of such a multipredictor analysis can provide rewarding insight and be used to advance the basis for regionalization of groundwater dynamics.

Data Availability Statement

Calculated index values for the time series and predictors can be found at Zenodo (<https://doi.org/10.5281/zenodo.3483451>). Groundwater time series cannot be provided publicly by the authors based on the data usage agreement with the LfU but can be downloaded online (from <https://www.gkd.bayern.de/en/groundwater/upper-layer> and <https://www.gkd.bayern.de/en/groundwater/deeper-layer>). A list of named observation wells can be found in the supporting information.

Acknowledgments

The authors would like to thank the German federal state agency Bayerisches Landesamt für Umwelt (LfU, <https://www.lfu.bayern.de>) for the provision of data and supporting information. We are thankful to Inge de Graaf, who guided the implementation of the DEM-derived boundary descriptors. Additional thanks go to the Rosa-Luxemburg Foundation, who funded the third author's scholarship. The work also partly contributes to the Formas project 2016-00513.

References

- Allen, D. M., Whitfield, P. H., & Werner, A. (2010). Groundwater level responses in temperate mountainous terrain: Regime classification, and linkages to climate and streamflow. *Hydrological Processes*, 24(23), 3392–3412. <https://doi.org/10.1002/hyp.7757>
- Barthel, R. (2011). An indicator approach to assessing and predicting the quantitative state of groundwater bodies on the regional scale with a special focus on the impacts of climate change. *Hydrogeology Journal*, 19(3), 525–546. <https://doi.org/10.1007/s10040-010-0693-y>
- Barthel, R. (2014). A call for more fundamental science in regional hydrogeology. *Hydrogeology Journal*, 22, 507–510. <https://doi.org/10.1007/s10040-014-1101-9>
- Beale, E. M. L., Kendall, M. G., & Mann, D. W. (1967). The discarding of variables in multivariate analysis. *Biometrika*, 54(3–4), 357–366. <https://doi.org/10.1093/biomet/54.3-4.357>
- BGR (2018). BÜK200 CC8726 V1.0, edited, Hannover.
- BGR (2019). Mittlere jährliche Grundwasserneubildung von Deutschland 1:1.000.000 (GWN1000). Digital map data v1, Hannover.
- Bloomfield, J. P., Allen, D. J., & Griffiths, K. J. (2009). Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. *Journal of Hydrology*, 373(1–2), 164–176. <https://doi.org/10.1016/j.jhydrol.2009.04.025>
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139235761>
- Boutt, D. F. (2017). Assessing hydrogeologic controls on dynamic groundwater storage using long-term instrumental records of water table levels. *Hydrological Processes*, 31, 1479–1497. <https://doi.org/10.1002/hyp.11119>
- Buffin-Bélanger, T., Cloutier, C.-A., Tremblay, C., Chaillou, G., & Larocque, M. (2015). Dynamics of groundwater floodwaves and groundwater flood events in an alluvial aquifer. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 41, 469–483. <https://doi.org/10.1080/07011784.2015.1102651>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., et al. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8, 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Cuthbert, M. O. (2014). Straight thinking about groundwater recession. *Water Resources Research*, 50, 2407–2424. <https://doi.org/10.1002/2013WR014060>
- de Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., & Bierkens, M. F. P. (2015). A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences*, 19, 823–837. <https://doi.org/10.5194/hess-19-823-2015>
- de Marsily, G., Delay, F., Gonçalves, J., Renard, P., Teles, V., & Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1), 161–183. <https://doi.org/10.1007/s10040-004-0432-3>
- Dettinger, M. D., & Diaz, H. F. (2000). Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology*, 1(4), 289–310. [https://doi.org/10.1175/1525-7541\(2000\)001<0289:GCOSFS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2)
- Eltahir, E. A. B., & Yeh, P. J.-F. (1999). On the asymmetric response of aquifer water level to floods and droughts in Illinois. *Water Resources Research*, 35(4), 1199–1217. <https://doi.org/10.1029/1998WR900071>
- Falkenmark, M., Chapman, T. G., & UNESCO (1989). *Comparative hydrology: An ecological approach to land and water resources*. Paris: UNESCO.
- Fetter, C. W. (1994). *Applied hydrogeology* (3rd ed.). New York: Macmillan College Publishing Company.
- Freeze, R. Allan, & Witherspoon, P. A. (1967). Theoretical analysis of regional groundwater flow: 2. Effect of water-table configuration and subsurface permeability variation. *Water Resources Research*, 3(2), 623–634. <https://doi.org/10.1029/WR003i002p00623>
- Giese, M., Haaf, E., Heudorfer, B., & Barthel, R. (2020). Comparative hydrogeology—Reference analysis of groundwater dynamics from neighbouring observation wells. *Hydrological Sciences Journal*, 65, 1685–1706. <https://doi.org/10.1080/02626667.2020.1762888>

- Gleeson, T., & Manning, A. H. (2008). Regional groundwater flow in mountainous terrain: Three-dimensional simulations of topographic and hydrogeologic controls. *Water Resources Research*, 44, W10403. <https://doi.org/10.1029/2008WR006848>
- G'Sell, M. G., Wager, S., Chouldechova, A., & Tibshirani, R. (2016). Sequential selection procedures and false discovery rate control. *Journal Of The Royal Statistical Society*, 78, 423–444. <https://doi.org/10.1111/rssb.12122>
- Haaf, E., & Barthel, R. (2018). An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs. *Journal of Hydrology*, 559, 222–237. <https://doi.org/10.1016/j.jhydrol.2018.02.035>
- Haitjema, H. M., & Mitchell-Bruker, S. (2005). Are water tables a subdued replica of the topography? *Groundwater*, 43(6), 781–786.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113, D20119. <https://doi.org/10.1029/2008JD010201>
- Heudorfer, B., Haaf, E., Stahl, K., & Barthel, R. (2019). Index-based characterization and quantification of groundwater dynamics. *Water Resources Research*, 55, 5575–5592. <https://doi.org/10.1029/2018WR024418>
- Heudorfer, B., & Stahl, K. (2017). Comparison of different threshold level methods for drought propagation analysis in Germany. *Hydrology Research*, 48, 1311–1326. <https://doi.org/10.2166/nh.2016.258>
- Knutsson, G., & Fagerlind, T. (1977). *Grundvattentillgångar i Sverige, Rapporter och meddelanden / Sveriges geologiska undersökning*, 0349-2176(9-). Stockholm: SGU.
- Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21, 2863–2879. <https://doi.org/10.5194/hess-21-2863-2017>
- Lafare, A. E. A., Peach, D. W., & Hughes, A. G. (2015). Use of seasonal trend decomposition to understand groundwater behaviour in the Permo-Triassic Sandstone aquifer, Eden Valley, UK. *Hydrogeology Journal*, 24, 141–158. <https://doi.org/10.1007/s10040-015-1309-3>
- Larkin, R. G., & Sharp, J. J. M. (1992). On the relationship between river-basin geomorphology, aquifer hydraulics, and ground-water flow direction in alluvial aquifers. *GSA Bulletin*, 104(12), 1608–1620. [https://doi.org/10.1130/0016-7606\(1992\)104<1608:OTRBRB>2.3.CO;2](https://doi.org/10.1130/0016-7606(1992)104<1608:OTRBRB>2.3.CO;2)
- Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10), 93–94. <https://doi.org/10.1029/2008EO100001>
- Machiwal, D., & Singh, P. K. (2015). Understanding factors influencing groundwater levels in hard-rock aquifer systems by using multi-variate statistical techniques. *Environmental Earth Sciences*, 74, 5639–5652. <https://doi.org/10.1007/s12665-015-4578-1>
- Marani, M., Eltahir, E., & Rinaldo, A. (2001). Geomorphic controls on regional base flow. *Water Resources Research*, 37(10), 2619–2630. <https://doi.org/10.1029/2000WR000119>
- Markovich, K. H., Manning, A. H., Condon, L. E., & McIntosh, J. C. (2019). Mountain-block recharge: A review of current understanding. *Water Resources Research*, 55, 8278–8304. <https://doi.org/10.1029/2019wr025676>
- QGIS Development Team (2018). QGIS: geographic information system. Open Source Geospatial Foundation Project. Retrieved from <http://qgis.osgeo.org>
- Rinderer, M., McGlynn, B. L., & van Meerveld, H. J. (2017). Groundwater similarity across a watershed derived from time-warped and flow-corrected time series. *Water Resources Research*, 53, 3921–3940. <https://doi.org/10.1002/2016WR019856>
- Rinderer, M., Meerveld, H. J., & McGlynn, B. L. (2019). From points to patterns: Using groundwater time series clustering to investigate subsurface hydrological connectivity and runoff source area dynamics. *Water Resources Research*, 55, 5784–5806. <https://doi.org/10.1029/2018WR023886>
- Rinderer, M., van Meerveld, I., Stähli, M., & Seibert, J. (2016). Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall? *Hydrological Processes*, 30, 1036–1051. <https://doi.org/10.1002/hyp.10634>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. *Water Resources Research*, 56, e2019WR025924. <https://doi.org/10.1029/2019wr025924>
- Seibert, J., Bishop, K., Rodhe, A., & McDonnell, J. J. (2003). Groundwater dynamics along a hillslope: A test of the steady state hypothesis. *Water Resources Research*, 39(1), 1014. <https://doi.org/10.1029/2002WR001404>
- Sklash, M. G., & Farvolden, R. N. (1979). The role of groundwater in storm runoff. *Journal of Hydrology*, 43(1–4), 45–65. [https://doi.org/10.1016/0022-1694\(79\)90164-1](https://doi.org/10.1016/0022-1694(79)90164-1)
- Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 7629–7634. <https://doi.org/10.1073/pnas.1507583112>
- Tibshirani, R., Tibshirani, R., Taylor, J., Loftus, J., & Reid, S. (2017). selectiveInference: Tools for Post-Selection Inference, edited.
- Tibshirani, R. J., Taylor, J., Lockhart, R., & Tibshirani, R. (2016). Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association*, 111, 600–620. <https://doi.org/10.1080/01621459.2015.1108848>
- Tiedeman, C. R., Goode, D. J., & Hsieh, P. A. (1998). Characterizing a ground water basin in a new England mountain and valley terrain. *Ground Water*, 36(4), 611–620. <https://doi.org/10.1111/j.1745-6584.1998.tb02835.x>
- Tóth, J. (1963). A theoretical analysis of groundwater flow in small drainage basins. *Journal of Geophysical Research*, 68(16), 4795–4812. <https://doi.org/10.1029/JZ068i016p04795>
- Tóth, J. (1970). A conceptual model of the groundwater regime and the hydrogeologic environment. *Journal of Hydrology*, 10(2), 164–176. [https://doi.org/10.1016/0022-1694\(70\)90186-1](https://doi.org/10.1016/0022-1694(70)90186-1)
- Vivoni, E. R., Di Benedetto, F., Grimaldi, S., & Eltahir, E. A. B. (2008). Hypsometric control on surface and subsurface runoff. *Water Resources Research*, 44, W12502. <https://doi.org/10.1029/2008WR006931>
- Vogt, J., Soille, P., de Jager, A., Rimavičiūtė, E., Mehl, W., Foisneau, S., et al. (2007). *A pan-European river and catchment database*. Luxembourg: OPOCE.
- Voltz, T., Gooseff, M., Ward, A. S., Singha, K., Fitzgerald, M., & Wagener, T. (2013). Riparian hydraulic gradient and stream-groundwater exchange dynamics in steep headwater valleys. *Journal of Geophysical Research: Earth Surface*, 118, 953–969. <https://doi.org/10.1002/jgrf.20074>
- Wilson, J. L., & Guan, H. (2004). Mountain block hydrology and mountain-front recharge. In J. F. Hogan, F. M. Phillips, B. R. Scanlon (Eds.), *Groundwater recharge in a desert environment: The southwestern United States* (pp. 113–117). Washington, DC: American Geophysical Union.
- Winter, T. C. (1978). Numerical simulation of steady state three-dimensional groundwater flow near lakes. *Water Resources Research*, 14(2), 245–254. <https://doi.org/10.1029/WR014i002p00245>

- Winter, T. C. (2001). The concept of hydrologic landscapes. *Journal of the American Water Resources Association*, 37(2), 335–349. <https://doi.org/10.1111/j.1752-1688.2001.tb00973.x>
- Winter, T. C., Mallory, S. E., Allen, T. R., & Rosenberry, D. O. (2000). The use of principal component analysis for interpreting ground water hydrographs. *Ground Water*, 38(2), 234–246. <https://doi.org/10.1111/j.1745-6584.2000.tb00335.x>
- Zecharias, Y. B., & Brutsaert, W. (1988). The influence of basin morphology on groundwater outflow. *Water Resources Research*, 24(10), 1645–1650. <https://doi.org/10.1029/WR024i010p01645>